

BAB II

TINJAUAN PUSTAKA

A. Penelitian Terdahulu

Penelitian terdahulu bertujuan untuk mendapatkan bahan perbandingan dan acuan. Selain itu, untuk menghindari anggapan kesamaan dengan penelitian ini. Maka dalam tinjauan pustaka, peneliti mencantumkan hasil-hasil penelitian terdahulu yang ditunjukkan pada Tabel 2.1.

Tabel 2. 1 Penelitian Terdahulu

No	Peneliti	Metode	Hasil
1	Artanti <i>et al.</i> (2018)	<i>Naïve Bayes Classifier</i>	Hasil analisis dari algoritma <i>Naïve Bayes</i> didapatkan tingkat rata-rata akurasi sebesar 93.33%.
2	Septiani (2022)	<i>Naïve Bayes Classifier</i>	Hasil dari analisis sentimen terhadap Go-Pay dan OVO menggunakan algoritma <i>Naïve Bayes</i> mendapatkan akurasi sebesar 82.81%.
3	Petiwi <i>et al.</i> (2022)	<i>Naïve Bayes Classifier</i> dan <i>Support Vector Machine</i>	Akurasi yang didapatkan pada penelitian ini yaitu <i>Support Vector Machine</i> mendapatkan hasil sebesar 83% Sedangkan akurasi <i>Naïve Bayes</i> sebesar 74,6%.
4	Putri <i>et al.</i> (2022)	<i>Naïve Bayes Classifier</i>	Hasil dari penelitian menggunakan algoritma <i>Naïve Bayes</i> didapatkan <i>accucary score</i> sebesar 80%
5	Susianti <i>et al.</i> (2020)	<i>Naïve Bayes Classifier</i>	Hasil dari penelitian ini adalah <i>Naive Bayes Classifier</i> cukup baik sebagai algoritma klasifikasi untuk analisis sentimen, dengan hasil <i>Accuracy</i> sebesar 87%
6	Pratama <i>et al.</i> (2019)	<i>Support Vector Machine</i> dan <i>Naïve Bayes Classifier</i>	Hasil klasifikasi Akurasi <i>Support Vector Machine</i> mendapatkan hasil sebesar 86.96% Sedangkan akurasi <i>Naïve Bayes</i> berbeda tipis yaitu sebesar 86.48%.
7	Farisi <i>et al.</i> (2019)	<i>Naïve Bayes Classifier</i>	Hasil akurasi yang didapatkan rata-rata 96,1% dengan menggunakan metode <i>Naïve Bayes Classifier</i>

Tabel 2. 1 Penelitian Terdahulu (lanjutan)

No	Peneliti	Metode	Hasil
8	Ratnawati (2018)	<i>Naïve Bayes Classifier</i>	Hasil pengujian dan analisis yang telah dilakukan dengan menggunakan algoritma <i>Naïve Bayes</i> . Hasil didapatkan yaitu <i>Accuracy</i> = 90%, <i>Precision</i> = 92% <i>Recall</i> = 90% <i>f-measure</i> = 90%
9	Ramadhani & Wahyudin (2022)	<i>Naïve Bayes</i> dan K-NN	Algoritma <i>Naïve Bayes</i> mendapatkan tingkat keakurasian yang paling tinggi yaitu dengan nilai akurasi sebesar 90.71% dibandingkan dengan algoritma KNN yang hanya 74.78%
10	Aulia & Patriya (2019)	<i>Lexicon Based</i> dan <i>Naïve Bayes Classifier</i>	Hasil yang didapat yakni dengan <i>Lexicon Based</i> yaitu sebesar 64.49% dan <i>Naïve Bayes</i> sebesar 86.53%.

B. Landasan Teori

1. Analisis Sentimen

Analisis sentimen adalah jenis pengolahan bahasa alami (*Natural language processing*) untuk melacak suasana hati masyarakat mengenai produk atau topik tertentu. Analisis sentimen juga disebut tambang pendapat, dimana sistem mengumpulkan dan meneliti pendapat tentang produk yang dibuat dalam bentuk komentar, postingan, ulasan, maupun *tweet*. Analisis sentimen dapat diimplementasikan dalam beberapa hal, seperti untuk pemasaran dengan melihat keberhasilan iklan atau peluncuran produk baru, menentukan versi produk baru atau jasa yang populer dan bahkan mengidentifikasi demografi masyarakat terhadap suka atau tidak suka terhadap suatu kebijakan yang dibuat oleh suatu instansi (Yahyadi & Latifah, 2022).

2. Pertukaran Mahasiswa Merdeka

Pertukaran Mahasiswa Merdeka adalah program pertukaran mahasiswa selama satu semester dari satu perguruan tinggi ke perguruan tinggi lainnya dengan sistem alih kredit sebanyak 20 SKS. Melalui program ini memungkinkan mahasiswa untuk belajar di perguruan tinggi lain dan dapat mendapatkan kesempatan untuk mengambil mata kuliah diluar program studi sesuai dengan pilihan mahasiswa (Pae *et al.*, 2022). Mahasiswa yang akan mengikuti pertukaran luar negeri maupun dalam negeri harus diakui oleh Fakultas dan Program Studi karena berkaitan dengan SKS yang diambil, dan harus adanya perjanjian kerjasama antar sesama kampus dalam program pertukaran mahasiswa.

Adapun tujuan dari program pertukaran pelajar mahasiswa ini menurut Fuadi (2021) antara lain:

1. Belajar lintas kampus (dalam dan luar negeri), sehingga terbangun persaudaraan lintas budaya dan suku.
2. Membangun persahabatan mahasiswa antar daerah, suku, budaya, dan agama sehingga terbangun semangat persatuan dan kesatuan bangsa.
3. Melaksanakan transfer ilmu pengetahuan untuk menutupi disparitas Pendidikan, baik antar perguruan tinggi dalam negeri maupun kondisi pendidikan tinggi dalam negeri dengan luar negeri.

3. *Text Mining*

Text mining merupakan proses pengelolaan data berupa teks yang dijadikan pola untuk mendapatkan sebuah informasi penting. Dalam *text mining* data yang diolah adalah data berupa tekstual dan bersifat abstrak. Sebelum melakukan *text mining*, terdapat beberapa tahap *text preprocessing*. Tahapan *text preprocessing* diantaranya yaitu *cleansing*, *case folding*, *tokenizing*, normalisasi, *filtering*, dan *stemming* (Pratama *et al.*, 2021).

4. Klasifikasi Teks

Klasifikasi teks merupakan salah satu penerapan dari *text mining*. Klasifikasi teks adalah proses pengelompokan suatu teks atau dokumen berdasarkan kata, frase, atau kombinasinya untuk menentukan kategori yang telah ditetapkan sebelumnya (*supervised learning*) (Sabrani *et al.*, 2020).

5. *Twitter*

Twitter adalah layanan jejaring sosial dan mikroblog yang memungkinkan penggunanya untuk mengirim, menerima dan membaca pesan berbasis teks hingga 140 karakter yang dikenal dengan sebutan kicauan (*tweet*). Dengan media sosial ini pengguna dapat mengekspresikan banyak pendapat, informasi, dan kegiatan sehari-hari.

Twitter merupakan salah satu jejaring sosial yang populer digunakan masyarakat di Indonesia. Menurut laporan Statista, terdapat 18,45 juta pengguna aplikasi yang didirikan oleh Jack Dorsey ini di

Tanah Air per Januari 2022 lalu. Capaian ini menempatkan Indonesia sebagai negara pengguna *Twitter* terbanyak ke-5 di dunia.

6. *Scraping*

Scraping data (web scraping) adalah proses mengekstraksi dan mengumpulkan data yang didapat dari sebuah situs *website* ke format yang baru (Nurmawati *et al.*, 2020). Dari data inilah akan dilakukan analisis dengan mengacu pada nilai sentimen data yang ada.

7. *Google Colaboratory*

Google Colaboratory (juga dikenal sebagai *Colab*) adalah *tools* yang dikembangkan oleh *Google Research* berbasis layanan *cloud* yang didasarkan pada *Jupyter Notebooks* untuk kebutuhan pendidikan dan penelitian *machine learning* (Fahmi *et al.*, 2020). *Google Colaboratory* dipilih dalam penelitian ini karena dapat digunakan secara *open source* dan terhubung langsung pada *google drive* sehingga mudah dipakai untuk menyimpan, menulis, serta membagikan program yang telah ditulis.

8. *Python*

Python merupakan bahasa pemrograman tingkat tinggi (*high level language*) yang dikembangkan oleh Guido van Rossum pada tahun 1989 dan diperkenalkan untuk pertama kalinya pada tahun 1991. *Python* lahir atas dasar keinginan untuk mempermudah seorang programmer dalam menyelesaikan tugasnya dengan cepat. *Python* memiliki beberapa kelebihan, yaitu pemrograman menggunakan *Python* jauh lebih cepat

dan lebih pendek dibandingkan menggunakan C++ dan Java, mampu menangani pemrograman kompleks dan mendukung pemrograman grafis serta platform *independent* yang berarti bahwa program yang dibuat menggunakan *Python* dapat berjalan pada sistem operasi apa saja selama terdapat platform *Python* (Phandany *et al.*, 2022).

9. *Lexicon Based Features*

Menurut Yahyadi dan Latifah (2022) *Lexicon Based Features* adalah kamus atau dataset yang berisikan kata-kata Bahasa Indonesia. Kamus *lexicon* digunakan untuk melakukan penilaian terhadap polaritas dari suatu *tweet*, dimana kata-kata akan di pasangkan dan dicocokkan nilai polaritasnya. Penggunaan *lexicon based features* dalam penelitian ini dimaksudkan untuk menentukan orientasi sentimen suatu kata.

10. *Naïve Bayes Classifier*

Naive Bayes Classifier merupakan sebuah metode klasifikasi menggunakan metode probabilitas dan statistik yg ditemukan oleh ilmuwan asal Inggris yang bernama Thomas *Bayes* pada abad ke-18. *Naive Bayes* memprediksi peluang di masa depan berdasarkan pengalaman di masa sebelumnya sehingga dikenal sebagai Teorema *Bayes* (Haqqizar & Larasyanti, 2019). Metode ini sering digunakan untuk menyelesaikan masalah dalam bidang mesin pembelajaran (*machine learning*) karena sederhana dan cepat dalam hal klasifikasi, kinerja yang baik dan mudah dalam penerapannya serta *Twitter*

memiliki tingkat akurasi yang tinggi dengan perhitungan sederhana (Sari et al., 2020).

11. *Confusion Matrix*

Menurut Nugroho (2018) *Confusion Matrix* adalah tool berguna untuk menganalisis seberapa baik algoritma klasifikasi atau *classifier* dalam mengenali kelas yang berbeda. *True Positive* (TP) dan *True Negative* (TN) menjelaskan ketika pengklasifikasi mendapatkan sesuatu dengan benar, sementara *False Positive* (FP) dan *False Negative* (FN) menjelaskan ketika pengklasifikasi mendapatkan hal yang salah. Dibawah ini menunjukkan Tabel *Confusion Matrix*.

Tabel 2. 2 *Confusion Matrix*

<i>Confusion Matrix</i>		<i>Predict Class</i>	
		Positif	Negatif
<i>Actual Class</i>	Positif	<i>TruePositif</i> (TP)	<i>FalsePositife</i> (FP)
	Negatif	<i>FalseNegative</i> (FN)	<i>TrueNegative</i> (TN)

Evaluasi dengan *Confusion Matrix* menghasilkan nilai *accucary*, *precision*, *recall* dan *F1-Score*. Berdasarkan Syarifuddin (2020) dijelaskan sebagai berikut:

Nilai *accucary* merupakan jumlah dokumen yang diklasifikasikan dengan benar, baik *TruePositive* maupun *TrueNegative*. Perhitungan nilai *accucary* dapat menggunakan persamaan 1.

$$accuracy = \frac{TP + TN}{TP + TN + FP + FN} \times 100\%$$

(1)

Nilai *precision* menggambarkan jumlah data kategori positif yang diklasifikasi secara benar dibagi dengan total data yang diklasifikasi positif. Dengan kata lain, *precision* adalah klasifikasi *TruePositive* dan semua data diprediksi sebagai kelas positif. Perhitungan nilai *precision* dapat menggunakan persamaan 2.

$$precision = \frac{TP}{(TP + FP)} \times 100\% \quad (2)$$

Nilai *recall* menunjukkan beberapa persen data kategori positif yang terklasifikasi dengan benar oleh sistem. Dengan kata lain, *recall* adalah jumlah dokumen yang memiliki klasifikasi *TruePositive* dari semua dokumen yang benar-benar positif (termasuk *FalseNegatif*). Perhitungan nilai *accuracy* dapat menggunakan persamaan 2.

$$recall = \frac{TP}{(TP + FN)} \times 100\% \quad (3)$$

F1-Score adalah hasil akhir dari perhitungan nilai *precision* dan *recall* untuk mencari rata-rata nilai *precision* dan *recall*. Perhitungan nilai *F1-Score* dapat menggunakan persamaan 4.

$$F1 - Score = \frac{2 \times Precision \times Recall}{Precision + Recall} \times 100\% \quad (4)$$